электронный журнал
МОЛОДЕЖНЫЙ НАУЧНО-ТЕХНИЧЕСКИЙ ВЕСТНИК
Издатель ФГБОУ ВПО "МГТУ им. Н.Э. Баумана". Эл No. ФС77-51038.

УДК 004.93

Использование информации о динамике изменений человеческого лица для решения задач распознавания и классификации

Сулимов А.С., *студент*
Россия, 105005, г. Москва, МГТУ им. Н.Э. Баумана,
кафедра «Программное обеспечение ЭВМ и информационные технологии»

Научный руководитель: Горин С.В., доцент
Россия, 105005, г. Москва, МГТУ им. Н.Э. Баумана
irudakov@bmstu.ru

Обработка человеческого лица в видеопотоке

Подходы к анализу человеческого лица в видеопотоке могут быть разделены на две группы, в зависимости от обрабатываемой временной информации. Первый подход основывается на обработке определенной последовательности кадров видеопотока, как отдельных статических изображений. Второй подход основывается на обработке динамических изменений структуры лица.

Антропологические исследования в работах [1, 3, 4, 11] показали, что динамические изменения человеческого лица при ведении разговора, а также движения головы представляют ключевую информацию для решения задач классификации (гендерной, возрастной). О'Тоул в исследовании [18], основываясь на физиологических особенностях человеческого лица и его изменении, выдвигает следующие принципы:

- и статическая, и динамическая информация могут быть использованы для решения задач распознавания;
- статическую информацию предпочтительно использовать для решения задач идентификации;
- динамическая информация позволяет получить качественный результат в условиях меняющегося окружения (освещение, разрешение изображений);
- модель, основанная на динамике изменений, требует большего времени на обучение;
- модель, основанная на динамике изменений, предпочтительна для решения задач гендерной классификации;

- для решения задач классификации эмоций динамическая информация является фундаментальной.

Ключевым моментом является понимание физиологии человеческой мимики. Выражение эмоций, например, счастья, грусти, страха, удивления, с точки зрения мимики заключается в определенном движении мышц лица, то есть изменении формы губ, век, кожи лица.

В рамках текущего исследования интерес представляет сравнение результатов распознавания базирующегося на статической и на динамической информации человеческого лица.

Для анализа статической информации человеческого лица и динамики его изменения авторы [14] предлагают использовать оператор локальных бинарных шаблонов, который результативно справляется с задачами распознавания объектов по шаблонам и применим к распознаванию динамических текстур.

Особенности обработки динамики изменения человеческого лица

Существуют различные подходы для обработки динамики изменения человеческого лица. В данной главе будут рассмотрены некоторые из них, наиболее подходящие для решения задач распознавания и классификации в видеопотоке.

В данной работе рассматриваются методы, которые основываются на анализе текстур изображений. Тукеран и Джейн провели классификацию методов данной группы, согласно которой выделили четыре группы: статистические, геометрические, основанные на моделях и обработке сигналов [12].

Раньше всех были предложены статистические методы [4]. С появлением различных трудов в сфере обработки сигналов, исследователи нашли их применение для обработки изображений. Так для распознавания человеческих лиц в видеопотоках Ли и Чен [16] предлагают использовать траектории отслеживаемых особенностей лица. Извлеченные с помощью фильтра Габора черты лица используются для составления модели распознавания. По результатам эксперимента авторы показали увеличение качества распознавания, основанного на описанной модели, в сравнении с обычным покадровым моделированием.

Из группы методов, основанных на построении модели обрабатываемого изображения, можно выделить работы, в которых описывается применение скрытых марковских моделей для решения задач распознавания человеческих лиц в видеопотоке. Ли и Чен в своей работе [7] приводят классификацию методов, основанных на данном подходе.

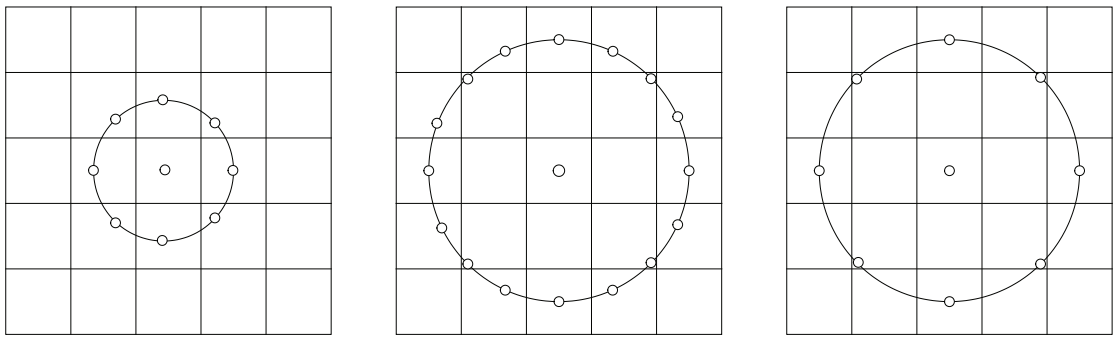
Большинство описанных подходов и методов не позволяют производить обработку и анализ текстур изображений в реальном времени из-за своей вычислительной сложности. Основным альтернативным подходом является использование локальных бинарных шаблонов (ЛБШ), впервые предложенных в 1996 году авторами Оджалой и Пьеткаиненом [9, 10]. Ос-

новным преимуществом является относительная малая сложность вычисления оператора ЛБШ. В связи с этим в последние годы были предложены новые методы для решения многих задач компьютерного зрения. Многие ученые разработали модификации оператора ЛБШ для конкретных задач, например, для работы с трехмерными текстурами [5], динамическими текстурами [12, 13].

Оператор локальных бинарных шаблонов для анализа статических свойств человеческого лица

Авторы ЛБШ, Оджала и Пьеткаинен, руководствовались идеей ассоциации каждого пикселя изображения с группой пикселей его окрестности [9, 10]. Применение оператора ЛБШ позволяет каждому пикселю полутонового изображения поставить в соответствие бинарный код, который описывает его текстурные характеристики.

Оператор работает с группой пикселей и вычисляет бинарный код для центрального пикселя группы. На рисунке 1 показаны фрагменты изображения размером 5x5 пикселей. Из рисунка 1 видно, что применение оператора ЛБШ зависит от количества пикселей окрестности, которыми описывается центральный пиксель области. На рисунке 1.а код центрального пикселя зависит от 8 соседних пикселей, на рисунке 1.б от 16-ти. Следует отметить, что соседние пиксели могут быть выбраны различными способами, на рисунке 1.в показано, как задать другие 8 соседних пикселей. То есть выбор «соседей» зависит также от их расстояния до целевого пикселя. В зависимости от конкретной задачи, качества изображения выбирается количество значимых пикселей, которое выбирается эмпирическим путем.



а) б) в)

Рис. 1. Группы пикселей для применения оператора ЛБШ

Каждый пиксель изображения имеет определенное значение интенсивности. Применение оператора ЛБШ позволяет вычислить бинарный код определенного пикселя, используя значения интенсивностей пикселей-соседей. Графическая иллюстрация применения оператора ЛБШ приведена на рисунке 2. Каждый квадрат условно описывает пиксель изображения. Так как оператор ЛБШ применим к полутоновым изображениям, то значения интенсивностей опреде-

ляются градациями серого в интервале $[0, 1]$, граничным значениям которого соответствуют 0 — белый цвет, 1 — черный цвет. Однако, обычно принято для удобства вычислений нормализовать значения интенсивности таким образом, чтобы значения интенсивности изменялись в интервале $[0, 100]$. На рисунке 2 значение интенсивности пикселя указано в центре квадрата. Координаты точек окрестности не всегда попадают точно в центры пикселей, поэтому для вычисления значений этих точек используется билинейная интерполяция.

Пиксели, которые имеют значения интенсивности больше, чем центральный пиксель (или равное ему), принимают значения «1», те, которые меньше центрального, принимают значения «0». Таким образом, получается бинарный код, представляющий окрестность пикселя.

Вычисление ЛБШ $\mathrm{LBP}_{\mathrm{P,R}}$ с радиусом R (на рисунке 1.а R=1, на рисунках 1.б, 1.в R=2) и количеством пикселей окрестности P производится последующей формуле (1):

$$\mathrm{LBP}_{\mathrm{P,R}}(\mathrm{x_c}, \mathrm{y_c}) = \sum_{\mathrm{p}=0}^{\mathrm{P}-1} \mathrm{s}\left(\mathrm{g_p} - \mathrm{g_c} + \mathrm{a}\right) \cdot 2^{\mathrm{p}}, \quad (1)$$

где $\mathrm{g_c}$ — значение интенсивности центрального пикселя $(\mathrm{x_c}, \mathrm{y_c})$ текущей области, $\mathrm{g_p}$ — p-ой точки окрестности. Для того, чтобы можно было регулировать работу оператора в зависимости от качества входного изображения вводится параметр a — специальное пороговое значение, и пороговая функция $\mathrm{s(x)}$ имеет вид (2):

$$\mathrm{s(x)} = \begin{cases} 1, \mathrm{x} \geq 0, \\ 0, \text{иначе.} \end{cases} \quad (2)$$

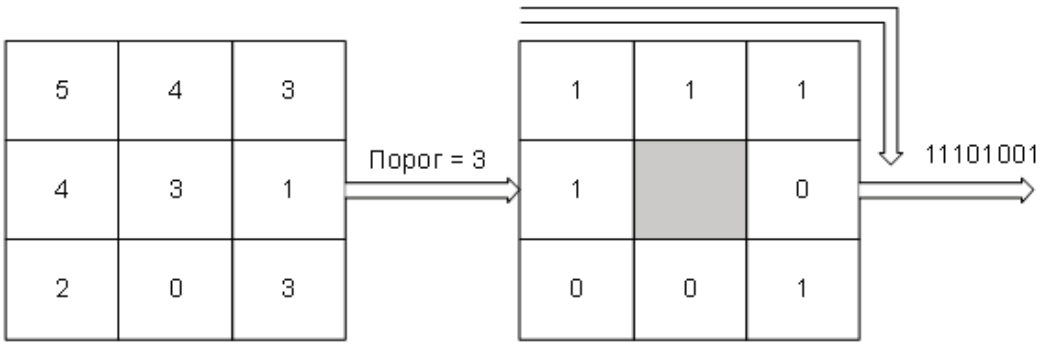


Рис. 2. Обработка с использованием ЛБШ

Из примера на рисунке 2 следует, что центральный пиксель описывается 8-миразрядным бинарным кодом 10010111_2. Стоит заметить, что выбор направления и начального пикселя для отсчета может быть произвольным. Так в примере на рисунке 2 был выбран левый верхний пиксель и направление обхода «соседей» по часовой стрелке. Согласно формуле (1), чтобы получить значение ЛБШ оператора необходимо привести полученный бинарный код к десятичной системе счисления, то есть получим:

$$10010111_2 = 1 + 2 + 4 + 16 + 128 = 151_{10}. \quad (3)$$

Вычисление гистограммы ЛБШ определяется по формуле (4).

$$H_i = \sum_{x,y} I(LBP(x,y) = i), \qquad i = 0, \dots, n-1,$$

где H − вычисляемая гистограмма, состоящая из n столбцов, n − максимальное значение бинарного кода, преобразованного в десятичную систему счисления. Величина n зависит от количества P учитываемых соседних пикселей $n = 2^P$. Функция I описывается формулой (4).

$$\mathrm{I(x)} = \begin{cases} 1, \mathrm{x} = \text{ИСТИНА} \\ 0, \mathrm{x} = \text{ЛОЖЬ}. \end{cases} \qquad (4)$$

Применение описанной методики позволяет отслеживать изменения не только каждого пикселя, но и его окрестности. Таким образом, представление заднего плана адаптируется к таким проблемам при обнаружении движения в видеопотоке как наличие шума, а также таких природных явлений как снег, дождь, качающаяся листва деревьев.

Оператор локальных бинарных шаблонов для анализа динамических свойств человеческого лица

Дзао в своей работе [16] предложил модификацию оператора ЛБШ для возможности его применения для анализа динамических текстур. Ученый предложил анализировать три идущих подряд кадра видеопоследовательности: текущий, предыдущий и последующий.

Произвольный пиксель кадра видеопотока описывается как $g_{0,c} = I(x, y, t)$ и определяется соответственно своими координатами (x, y) и моментом времени t. Таким образом целевые (центральные) пиксели описываются формулой (5).

$$g_{i,c} = I(x, y, t + i \cdot \Delta t), \qquad i = -1, 0, 1, \qquad (5)$$

где x, y − координаты пикселя, t − момент времени появления кадра в видеопоследовательности, Δt − промежуток времени между последовательными кадрами.

Соседние P пикселей выбираются аналогично статическому ЛБШ по формуле (6).

$$g_{i,p} = I(x + x_p, y + y_p, t + i \cdot \Delta t), \qquad p = 0, \dots, P-1; \quad i = -1, 0, 1, \qquad (6)$$

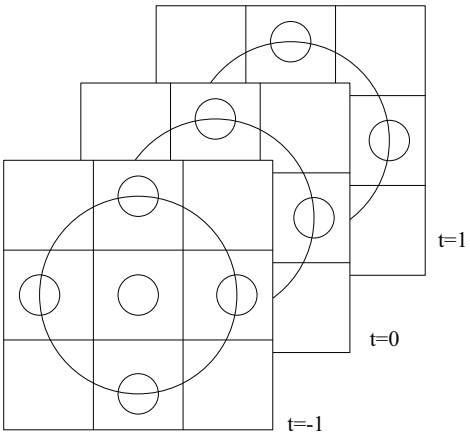
где P − количество соседних пикселей.

Динамический оператор ЛБШ, зависящий от промежутка времени между последовательными кадрами Δt, количеством соседних пикселей P и расстоянием между целевым и соседними пикселями R, вычисляется по формуле (7).

$$VLBP_{\Delta t,P,R} = \sum_{q=0}^{3P+1} \vartheta_q 2^q, \qquad (7)$$

где ϑ_q — определяет значение функции (2), аргументами которой являются разности вида $(g_{t,p} - g_{0,c})$, количество которых составляет $(3P + 2)$.

На рисунке 3 показана вычислительная процедура динамического ЛБШ, для $\Delta t = 1, P = 4, R = 1$. Первым шагом является получение последовательных кадров видеоизображения (см. рисунок 3.а). После того как получены кадры видеопотока происходит вычисление интенсивности целевых и соседних пикселей (см. рисунок 3.б). Следующим этапом является пороговая обработка в результате, которой значения интенсивностей соседних пикселей становятся равными «0» или «1». Заключительным шагом является вычисление бинарного кода динамического ЛБШ и его преобразование в десятичную систему счисления.



а)

	17	
31	30	99
	11	

	25	
35	27	88
	13	

	38	
31	33	77
	13	

б)

	0	
1	1	1
	0	

	0	
1		1
	0	

	1	
1	1	1
	0	

в)

Рис. 3. а) Иллюстрация последовательности кадров видеоизображения. б) Значения интенсивностей пикселей кадров. в) Значения интенсивностей пикселей каждого кадра после пороговой обработки

Результатом применения оператора ко всему изображению аналогично по формуле (4) является гистограмма, которая впоследствии может быть подана на вход классификатору для дальнейшей обработки.

Описание эксперимента

Размеченная база данных с тестовыми видеоизображениями

При составлении тестовой базы данных были установлены следующие признаки, которые необходимо учитывать при разработке метода:

- количество людей в сцене,
- ракурс, в котором человек движется в сцене,
- динамичный или статичный фон,
- проблемы, связанные с качеством изображений, разрешением, наличием шума.

Основными же признаками, по которым составляется база данных, являются:

- пол человека,
- возраст человека.

Требования, которыми руководствовались авторы для составления базы данных:

- разные сцены;
- разные положения людей (разные признаки).

При составлении базы, каждой записи ставится в соответствие вектор признаков, который описывает экспертную оценку видеопотока по описанным признакам.

Оценка результатов

В качестве меры оценки точности работы методов используется так называемая частота распознавания, которая рассчитывается по формуле (8).

$$E = \frac{N_P}{N}. \qquad (8)$$

где N_P — количество верно распознанных фрагментов, N — количество распознаваемых фрагментов.

Исходные данные и постановка задач

Для сравнения описанных подходов используются тестовые данные, которые предоставляются организациями, ведущими научные исследования в сфере компьютерного зрения. В зависимости от решаемых задач используются наборы данных:

- CRIM состоит из 591 видеофрагмента, размер кадров 130 x 150 пикселей [18];
- VidTIMIT включает в себя 720 видеоизображений, размер кадров 320 x 240 пикселей [19].

В качестве параметров алгоритма ЛБШ выбираются следующие:

- количество соседних пикселей $P = 8$;
- расстояние от центральных пикселей до соседних $R = 1$ пикс.

Для применения динамического ЛБШ также учитывается значение промежутка времени между последовательными кадрами $\Delta t = 0{,}5$с.

Значение параметров были подобраны эмпирическим путем при обработке одного случайным образом выбранного видеофрагмента из тестовых баз данных.

Используемые средства

В качестве среды для проведения экспериментов был выбран пакет прикладных программ MATLAB. Для выделения особых характеристических точек при анализе человеческих лиц в видеопотоке использовалась свободно распространяемая библиотека The Machine Perception Toolbox [14].

Для классификации используются средства библиотеки LIBSVM [8].

Описание эксперимента

В листинге (1) представлен код для получения гистограммы LBP. В листинге (2) представлен код для вычисления VLBP, основанного на информации целого фрагмента видеопотока. Исходные видеофрагменты помещаются в отдельные папки в зависимости от проводимого эксперимента. К каждому фрагменту применяется вычисление гистограммы LBP/VLBP, и генерируются файлы для обучения классификатора.

После этого к исходным видеофрагментам применяется оператор LBP, и полученная гистограмма подается на вход классификатору. В зависимости от результата классификации исходный видеофрагмент помещается в соответствующую папку. Зная имена исходных видеофрагментов, определяется количество корректно классифицированных данных.

Распознавание человеческих лиц

Исследовалась работа метода с видеоизображениями, имеющих различные размеры кадров 130x150 пикселей, содержащихся в CRIM [18], 320x240 пикселей из базы VidTIMIT [19]. Также производилось масштабирование видеофрагментов до размера кадра 40x30 пикселей.

В таблице 1 представлены результаты распознавания, полученные на описанной базе видеофрагментов. Из результатов видно, что метод, основанный на динамике изменений человеческого лица, значительно превосходит в частоте распознавания метод, использующий исключительно статические кадры для составления модели.

Стоит отметить, что при более высоком разрешении точность распознавания обоих методов выше. Однако при значительном размере кадра результаты изменяются незначительно.

Полученные результаты подтверждают выдвинутые в [11] принципы.

Таблица 1

Результаты распознавания

Метод	Результат, %		
	40x30 пкс.	130x150 пкс.	320x240 пкс.
Основанный на статике изображений	89,1	93,3	94,0
Основанный на динамике изменений изображений	94,7	98,1	98,2

Определение пола

В данном эксперименте в качестве используемых видеофрагментов для обучения и тестирования берутся изображения из свободно распространяемых баз данных CRIM [18], VidTIMIT [19].

В процессе предобработки в каждом входном видеофрагменте осуществляется выделение особых точек — глаз и по положению глаз вычисляется область лица. Так как в данном и последующем экспериментах входные данные берутся из разных баз и не унифицированы, необходимым является масштабирование полученных видеофрагментов.

Результаты классификации приведены в таблице 2.

Таблица 2

Результаты гендерной классификации

Метод	Результат, %		
	40x30 пкс.	130x150 пкс.	320x240 пкс.
Основанный на статике изображений	90,6	93,4	92,1
Основанный на динамике изменений изображений	80,1	89,2	94,7

Из полученных результатов следует, что использование динамической информации при данном подходе не позволяет улучшить производительность метода, при обработке изображений низкого разрешения. При увеличении размера кадра точность метода, основанного на применении VLBP, превосходит показатели статического метода.

Определение возраста

Как и в предыдущем примере в настоящем эксперименте используются свободно распространяемые базы видеофрагментов CRIM [18], VidTIMIT [19].

Процесс предобработки схож с описанным ранее и заключается в установлении особых точек, по которым определяется область лица, и масштабировании видеофрагментов до унифицированных размеров.

Для классификации были выбраны следующие возрастные группы:

- до 9 лет,
- от 10 до 19 лет,
- от 20 до 39 лет,
- от 40 до 59 лет,
- от 60 лет.

Результаты классификации приведены в таблице 3.

Таблица 3

Результаты возрастной классификации

Метод	Результат, %		
	40x30 пкс.	130x150 пкс.	320x240 пкс.
Основанный на статике изображений	77,6	83,2	83,9
Основанный на динамике изменений изображений	69,1	68,7	69,0

Полученные результаты показывают, что использование динамической информации не позволяет улучшить качество классификации человеческих лиц по возрасту.

Выводы

В настоящей работе были рассмотрены антропологические исследования в области физиологии человеческого лица и его применимости для решения задач распознавания. Также были выделены работы, в которых особое значение уделяется динамике изменения человеческого лица, мимике.

Для того, чтобы проследить достоинства и недостатки использования данного подхода, были проведены следующие эксперименты:

- распознавание человеческого лица в видеопотоке;
- гендерная классификация;
- возрастная классификация.

В каждом эксперименте проверялись два метода, использующих модели, основанные на:

- статических характеристиках человеческого лица;
- динамике изменений человеческого лица.

Также прослеживалась зависимость влияния размера кадров видеоизображения на точность распознавания.

В качестве алгоритма выделения характеристических особенностей к кадрам видеопотока применялся оператор LBP, для классификации использовался SVM.

Экспериментально было показано, что для решения задачи распознавания использовать информацию о динамики изменений человеческого лица целесообразно, что соответствует антропологическим исследованиям.

Также было установлено, что для решения задач классификации в рамках проводимых экспериментов целесообразно применять подход, основанный на динамической информации для видеоизображений размером от 320x240 пикселей. При меньшем разрешении применение данного метода не дает выигрыш в точности классификации.

В качестве дальнейших исследований необходимо рассмотреть и проанализировать методы определения характеристических точек, подходящих для решения широкого круга задач, от задач распознавания до классификации.

Листинги

Листинг 1. Получение гистограммы LBP для одного кадра видеопотока.

```
I = imread('test.png');
mapping = getmapping(8, 'u2');
H = LBP(I, 1, 8, mapping, 'h');
```

Листинг 2. Получение гистограммы VLBP для последовательности кадров видеопотока.

```
cd ('..\test\');
a = dir('*.jpg');
for i = 1 : length(a)
    img_name = getfield(a, {i}, 'name');
    img_dat = imread(img_name);
    % Конвертирование в оттенки серого
    if size(img_dat, 3) == 3
        img_dat = rgb2gray(img_dat);
    end
    [height width] = size(img_dat);
    if i == 1
        vol_data = zeros(height, width, length(a));
    end
    vol_data(:, :, i) = img_dat;
end
cd ..

rotate_index = 1;
radius = 1;
```

```
19    time_int = 0.5;
20    neighbors = 8;
21    time_len = 1;
22    border_len = 1;
23    bil_interpolation = 1;
24    H = RIVLBP(vol_data, time_int, radius,
25                neighbors, border_len,
26                time_len, rotate_index, bil_interpolation);
```

Список литературы

1. Bassili, J. Emotion recognition: The role of facial movement and the relative importance of upper and lower areas of the face / J. Bassili // Journal of Personality and Social Psychology. — 1979. — Vol. 37, no. 27(2). — P. 2049–2059.
2. Haralick, R.M., Dinstein, I., Shanmugaman, K. Textural features for image classification IEEE Trans. Syst. Man Cybern. SMC-3, 1993, 610-621.
3. Hill H., Johnson A. Categorizing sex and identity from the biological motion of faces / Johnson A. Hill, H. // Current Biology. — 2001. — no. 11(11). — P. 880–885.
4. Knight B., Johnston A. The role of movement in face recognition / Johnston A. Knight, B. // Visual Cognition. — 1997. — no. 4. — P. 265–274.
5. Leung, T., Malik, J.: Representing and recognizing the visual appearance of materials using theedimensional texons. Int. J. Comput. Vis. 43(1), 29-44, 2001.
6. LIBSVM URL: http://www.csie.ntu.edu.tw/ cjlin/libsvm/.
7. Liu X., Chen T. Video-Based Face Recognition Using Adaptive Hidden Markov Models / Liu X., Chen, T. // IEEE. 2003, Pp. 340-345.
8. Maenpaa, T. The Local Binary Pattern Approach To Texture Analysis – Extensions And Applications: Ph.D. thesis / Infotech Oulu and Department of Electrical and Information Engineering, University of Oulu. — 2003.
9. Ojala, T., Pietikäinen, M. and Harwood D. Performance evaluation of texture measures with classification based on Kullback discrimination of distributions, Proceedings of the 12th IAPR International Conference on Pattern Recognition, — 1994, vol. 1, pp. 582 - 585.
10. Ojala, T., Pietikäinen, M. and Harwood D. A Comparative Study of Texture Measures with Classification Based on Feature Distributions, Pattern Recognition, — 1996, vol. 29, pp. 51-59.
11. O'Toole A.J., Roark D.A. Abdi H. Recognizing moving faces: A psychological and neural synthesis / Roark D.A. Abdi H. O'Toole, A.J. // Trends in Cognitive Science. — 2002. — no. 6. — P. 261–266.

12. Saisan, P., Doretto, G., Wu, Y.N., Soatto, S. Dynamic texture recognition. In: Proc. IEEE Conference of Computer Vision and Pattern Recognition, vol. 2, pp. 5863, 2001.
13. Szummer, M., Jain, A.K. Temporal texture modeling. In: Proc. IEEE International Conference of Image Processing, vol 3, pp. 823826, 1996.
14. The Machine Perception Toolbox URL: http://mplab.ucsd.edu/grants/project1/free-software/MPTWebSite/introduction.html.
15. Tuceryan, M., Jain, A.K.: Texture Analysis. In: Chen, C.H., Pau, L.F., Wang, P.S., The Handbook of Pattern Recognition and Computer Vision, 2nd edn., pp. 207-248. World Scientific, Singapore(1998).
16. Zhao G., Pietikainen M. Dynamic texture recognition using local binary patterns with an application to facial expressions / Pietikainen M. Zhao, G. //IEEE TPAMI. — 2007. — Vol. 29(6). — Pp. 915–928.
17. Маслий, Р. В. Использование локальных бинарных шаблонов для распознавания лиц на полутоновых изображениях / Р. В. Маслий // Информационные технологии и компьютерная техника. — 2008. — Т. 4. — С. 6.
18. Наборы данных для алгоритмов распознавания речи и мимики лица [Электронный ресурс]. — URL: http://www.crim.ca/en, (дата обращения: 18.12.2012г.).
19. Наборы данных для алгоритмов распознавания речи и мимики лица [Электронный ресурс]. — URL: http://itee.uq.edu.au/ conrad/vidtimit/, (дата обращения: 18.12.2012г.).